RESUMEN DE ESTADÍSTICA

<u>TEMA 2:</u> DESCRIPCIÓN UNIVARIANTE

MEDIDAS DE POSICIÓN

Media aritmética

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{j} x_i n_i$$

Mediana, cuartiles, deciles y percentiles

Para calcular estos estadísticos, en primer lugar obtenemos el valor *z* entre 0 y 1 que representa a la posición que ocupa la mediana, cuartil, decil o percentil.

- Mediana: z = 1/2
- Cuartil: z = k/4, donde k es el nº de cuartil (1, 2 ó 3)
- Decil: z = k/10, donde k es el nº de decil (de 1 a 9)
- Percentil: z = k/100, donde k es el nº de percentil (de 1 a 99)

1. Ordenamos los datos de menor a mayor.
2. Nos fijaremos en la frecuencia acumulada para encontrar la mediana/cuartil/...
3. Si tenemos la frecuencia absoluta (n_i) pero no la relativa (f_i), multiplicamos z por N.
4. Nos fijamos en el primer valor cuya frecuencia sea igual o superior a z (o z*N).
5. Si el valor es superior a z (o z*N), el valor encontrado es la mediana/cuartil/…
6. Si el valor es igual a z (o z*N), la mediana/cuartil/… se obtiene calculando la media entre el valor encontrado y el inmediatamente superior.

<u>Ejemplos:</u>

x_i	n_i	N_i
10	4	4
11	6	10
12	7	17
13	3	20

x_i	f_i	F_i
10	0,1	0,1
11	0,2	0,3
12	0,3	0,6
13	0,4	1

1. En la tabla de la izquierda queremos calcular la mediana (Me).
   z = 1/2 → Como tenemos la frecuencia absoluta, z*N = (1/2)*20 = <u>10</u>.
   En el valor Xi=<u>11</u>, Ni=10 es igual a z*N. Por tanto, Me = (11+12)/2 = **11,5**
2. En la tabla de la derecha queremos calcular el 4º decil (D_4).
   z = 4/10 → En el valor Xi=<u>12</u>, Fi=0,6 supera a z. Por tanto, D_4 = **12**

Cálculo de la mediana/cuartil/… con datos agrupados en intervalos

Se siguen los mismos pasos que antes, pero una vez que tenemos el intervalo de la mediana/cuartil/…, hay que calcular el valor más apropiado dentro de ese intervalo.

Fr. relativas: $P_z = E_{i-1} + A_i\left(\dfrac{z - F_{i-1}}{f_i}\right)$ Fr. absolutas: $P_z = E_{i-1} + A_i\left(\dfrac{z \cdot N - N_{i-1}}{n_i}\right)$

- z es un valor entre 0 y 1 que representa la posición de la mediana/cuartil/...
  Por ejemplo, para el primer cuartil z=1/4 (ver página anterior).
- E_{i-1} es el extremo inferior del intervalo donde está la mediana/cuartil/…
- A_i es la amplitud de ese intervalo.
- F_{i-1} (o N_{i-1}) es la frecuencia acumulada del intervalo anterior.
- f_i (o n_i) es la frecuencia del intervalo donde está la mediana/cuartil/…
- Elegiremos una u otra fórmula según tengamos frecuencias relativas o absolutas.

Ejemplo:

E_{i-1}	E_i	n_i	N_i
10	15	3	3
20	25	8	11
30	35	6	17
35	40	3	20

Queremos calcular el 3º cuartil. En este caso: z = 3/4

Como tenemos frecuencias absolutas, calculamos z*N = 3/4*20=15.

Buscamos el intervalo cuya frecuencia absoluta acumulada sea mayor que 15.

Por tanto, el 3º cuartil se encuentra en el intervalo de 30 a 35.

$$P_{3/4} = E_{i-1} + A_i\left(\frac{z \cdot N - N_{i-1}}{n_i}\right) = 30 + 5\left(\frac{(3/4)20 - 11}{6}\right) = 30 + 5 \cdot \frac{2}{3} = 33,\widehat{3}$$

Propiedades de la media

1. $\sum_{i=1}^{j}\left(X_i - \overline{X}\right)n_i = 0$
2. Si $Z_i = C + X_i \quad \Rightarrow \quad \overline{Z} = C + \overline{X}$
3. Si $Z_i = C \times X_i \quad \Rightarrow \quad \overline{Z} = C \times \overline{X}$

Moda

La moda es el valor que más se repite en una distribución.

Si los datos están agrupados en intervalos de igual amplitud:

Localizamos el intervalo que más se repite (intervalo modal) y usamos esta fórmula:

$$Mo = E_{i-1} + A_i \left(\frac{n_{i+1}}{n_{i-1} + n_{i+1}} \right)$$

- E_{i-1} es el extremo inferior del intervalo modal.
- A_i es la amplitud del intervalo modal.
- n_i es la frecuencia del intervalo modal.
- n_{i-1} es la frecuencia del intervalo anterior al modal.
- n_{i+1} es la frecuencia del intervalo posterior al modal.
- Pueden usarse frecuencias relativas o absolutas.

Si los datos están agrupados en intervalos de distinta amplitud:

Construimos la columna d_i, dividiendo las frecuencias entre cada amplitud: $d_i = n_i / A_i$

Luego realizamos los mismos pasos que en el caso anterior, pero fijándonos en d_i en lugar de n_i. Es decir, localizamos el intervalo con mayor d_i y aplicamos esta fórmula:

$$Mo = E_{i-1} + A_i \left(\frac{d_{i+1}}{d_{i-1} + d_{i+1}} \right)$$

MEDIDAS DE DISPERSIÓN

Desviación media

$$D_m = \frac{1}{N} \sum_{i=1}^{j} \left| x_i - \overline{x} \right| n_i$$

Varianza

$$\sigma_x^2 = \frac{1}{N} \sum_{i=1}^{j} \left(x_i - \overline{x} \right)^2 n_i = \left(\frac{1}{N} \sum_{i=1}^{j} x_i^2 n_i \right) - \overline{x}^2$$

Desviación típica

$$\sigma_x = \sqrt{\sigma_x^2}$$

Propiedades de la varianza y la desviación típica

Si $Z_i = C + X_i \quad \Rightarrow \quad Var(Z) = Var(X) \quad \text{y} \quad \sigma(Z) = \sigma(X)$

Si $Z_i = C \times X_i \quad \Rightarrow \quad Var(Z) = C^2 \times Var(X) \quad \text{y} \quad \sigma(Z) = C \times \sigma(X)$

Coeficiente de variación (de Pearson)

Es una medida normalizada de la dispersión de una distribución.

$$CV = \frac{\sigma}{\overline{x}}$$

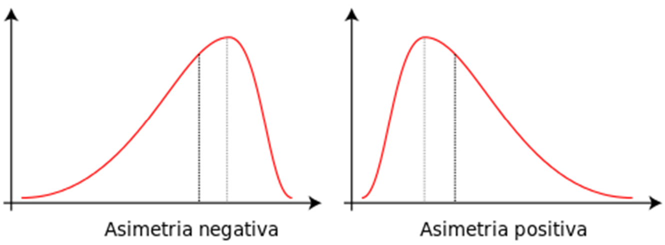
- Si CV < 1, la distribución se considera de “poca variabilidad”.
- Si CV > 1, la distribución se considera de “alta variabilidad”.
- Si CV > 1, la media es poco representativa (y viceversa).

MEDIDAS DE FORMA

Coeficiente de asimetría de Fisher

$$CA = \frac{\mu^3}{\sigma^3} = \frac{\frac{1}{N}\sum_{i=1}^{j}\left(x_i - \overline{x}\right)^3 n_i}{\sigma^3}$$

- Si CA < 0 hay asimetría a la izquierda (asimetría negativa).
- Si CA = 0 hay simetría.
- Si CA > 0 hay asimetría a la derecha (asimetría positiva).

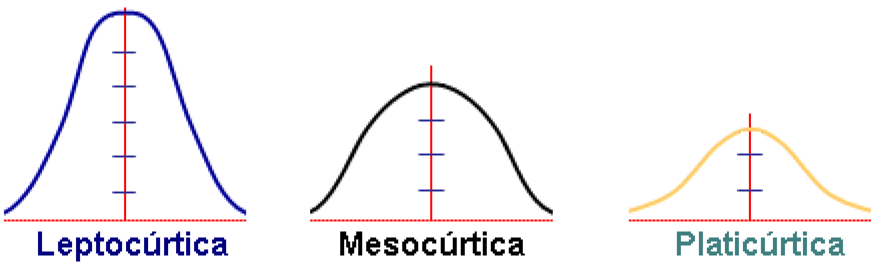


Coeficiente de curtosis

$$K = \frac{\mu^4}{\sigma^4} = \frac{\frac{1}{N}\sum_{i=1}^{j}\left(x_i - \overline{x}\right)^4 n_i}{\sigma^4}$$

- Si K = 3 la distribución es mesocúrtica.
- Si K > 3 la distribución es leptocúrtica.
- Si K < 3 la distribución es platicúrtica.

Nota: A veces se le resta 3 al coeficiente de curtosis para “normalizarlo”.



MEDIDAS DE CONCENTRACIÓN

Índice de Gini

$$I_G = \frac{\sum_{i=1}^{j-1} (p_i - q_i)}{\sum_{i=1}^{j-1} p_i} \qquad p_i = \frac{N_i}{N} \times 100 \qquad q_i = \frac{u_i}{u_n} \times 100$$

Para cada elemento, u_i es la suma acumulada de x_i*n_i en ese elemento:

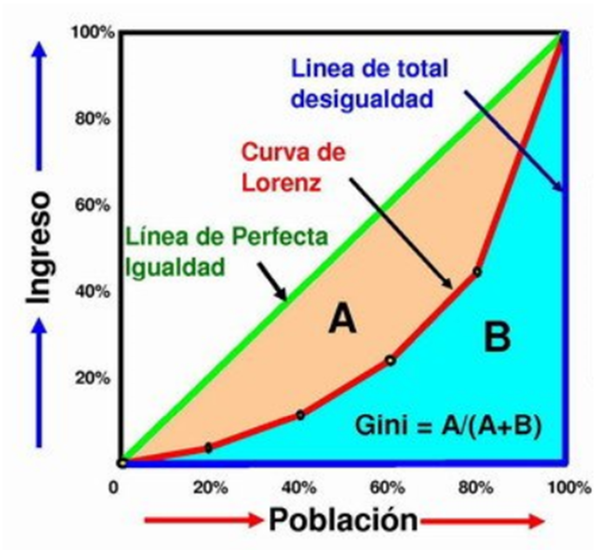
$$u_i = x_1 n_1 + x_2 n_2 + \ldots + x_i n_i = \sum_{h=1}^{i} x_h n_h$$

<u>Ejemplos:</u> práctica 5; http://youtu.be/STSi41E1YUY

Curva de Lorenz

Se construye a partir de los valores de p_i y q_i que se calculan para el índice de Gini.

El área hasta la bisectriz representa el grado de concentración de la variable.



<u>TEMAS 3-4:</u> DESCRIPCIÓN BIVARIANTE. REGRESIÓN

OMITIDO.
(Aprende las fórmulas y practica con los ejercicios.)

TEMA 5: SERIES TEMPORALES

CONSEJO: Para este tema es conveniente repasar la práctica 8.

Introducción

Serie temporal: Sucesión de observaciones cuantitativas ordenadas cronológicamente (o distribución bidimensional, con X=tiempo, Y=variable).

En este tema aprenderemos fundamentalmente a describir el comportamiento de las series temporales y a predecir la serie en el futuro.

Componentes de una serie

- Tendencia (T): Refleja la evolución a largo plazo.
- Estacionalidad (E): Recoge las oscilaciones que se producen en períodos < 1 año y que se repiten de forma regular (ej.: por vacaciones, Navidad, etc.).
- Ciclo (C): Recoge las oscilaciones periódicas de amplitud mayor a 1 año (ej.: etapas de prosperidad o depresión).
- Componente residual (R): Recoge fluctuaciones erráticas debidas a fenómenos imprevisibles.

Nosotros nos centraremos en el cálculo de las componentes tendencial y estacional.

Esquema aditivo o multiplicativo

La serie puede tener un esquema aditivo o multiplicativo según se sumen o multipliquen sus componentes:

- Esquema aditivo: $Y_t = T_t + E_t + C_t + R_t$
- Esquema multiplicativo: $Y_t = T_t * E_t * C_t * R_t$

Cuando en un ejercicio no se indica el esquema, supondremos que es aditivo.

¿Cómo detectar si la serie tiene esquema aditivo o multiplicativo?

1. Calcular medias y desviaciones típicas para cada año.
2. Representar los valores en un eje de coordenadas (X=media; Y=desviación típica).
3. Analizar la nube de puntos: creciente → multiplicativo; otra cosa → aditivo.

CÁLCULO DE LA TENDENCIA (T_t)

Método de medias móviles

(Ver ejemplo en práctica 8.)

1. Se determina el valor de *p* (nº de períodos en un año).
2. Se calcula la media de los p-primeros valores y se coloca en la casilla que esté en el centro de esos valores.
3. Repetimos el paso anterior avanzando un valor en cada vez.

Medias móviles cuando *p* es par

Si *p* es par, primero se calculan las medias móviles en una columna auxiliar y luego se obtiene la tendencia en otra columna, calculando las medias de 2 en 2. Ejemplo:

t	Yt	Media móvil	Tendencia (Tt)	Yt-Tt
1997.1	78.92			
1997.2	79.19	79.62		
1997.3	80.05	80.09	79.86	0.19
1997.4	80.32	80.60	80.35	-0.03
1998.1	80.82	81.06	80.83	-0.01
1998.2	81.21	81.62	81.34	-0.13
1998.3	81.88			
1998.4	82.58			

Serie sin tendencia

A veces es necesario calcular la serie sin tendencia, como en la última columna de la tabla anterior. Para ello, a cada valor se le resta su tendencia, si el esquema es aditivo, o se le divide, si el esquema es multiplicativo.

Método analítico

Primero se obtiene la recta de regresión para la serie, siendo X el tiempo e Y la variable (para X inventamos valores consecutivos; ej. 1, 2, 3…).

$$\hat{y}_t = \alpha + \beta x_t \qquad \beta = \frac{\mathrm{Cov}(x, y)}{\mathrm{Var}(x)} \qquad \alpha = \overline{y} - \beta \overline{x}$$

Luego sustituimos cada valor de X en la recta obtenida, para calcular la tendencia.

Ejemplo: Puedes descargar un ejemplo del método analítico en esta dirección:
www.jaimedv.com/eco/1c1-est/practica-8-metodo-analitico.xlsx

Este ejemplo también incluye el cálculo de la componente estacional.

COMPONENTE ESTACIONAL (E_t)

Calcularemos la componente estacional de cada período dentro del año. Para ello:

1. Hallamos la media de la serie sin tendencia de cada estación (Mi).
2. Calculamos la media de las medias anteriores (media anual, MA).
3. A cada media estacional le restamos (o dividimos, si es esq. mult.) la media anual.
   Adit.: $E_t = M_i - MA$ Mult.: $E_t = M_i / MA$

PREDICCIÓN

A partir de las componentes tendencial y estacional, podemos obtener la predicción para cualquier valor de *t*. Para ello sumamos o dividimos las componentes:

- Esquema aditivo: $Y_t = T_t + E_t$
- Esquema multiplicativo: $Y_t = T_t * E_t$

Al aplicar la fórmula tendremos en cuenta que:

- T_t es la componente tendencial, que se calcula con la recta de regresión del método analítico.
- E_t es la componente estacional. Cogeremos la que corresponda a la estación que queremos predecir.

ELIMINACIÓN DIRECTA DE LAS COMPONENTES

En este apartado veremos cómo eliminar las componentes sin tener que calcularlas.

Eliminación de la tendencia por diferencias regulares

Se construye una serie nueva restando el valor actual menos el anterior.

t	Y_t	Z_t
1992.1	3319	
1992.2	3929	610
1992.3	4154	225
1992.4	3907	–247
1992.5	3040	–867
1992.6	4938	1898

Eliminación de la estacionalidad por diferencias estacionales

Se resta el valor actual menos el de la misma estación del año anterior.

t	Y_t	Z_t
1992.1	80.05	
1992.2	80.32	
1992.3	80.82	
1993.1	81.21	1,16
1993.2	81.88	1,56
1993.3	82.58	1,76

TASA REGULAR DE VARIACÓN

(Ver ejemplo en práctica 9.)

Miden los cambios que se producen en cualquier serie temporal. 3 tipos:

- T.V. regular: variación entre 2 observaciones consecutivas.
- T.V. estacional: variación entre la misma estación de 2 años consecutivos.
- T.V. acumulada: variación en los r primeros meses del año, con respecto al mismo período del año anterior.

<u>TEMA 6:</u> NÚMEROS ÍNDICES

CONSEJO: Para este tema es conveniente repasar las prácticas 9 y 10, así como el archivo “Ejemplo Número Índices.xls”.

Introducción: Los números índices estudian la <u>evolución temporal</u> de variables estadísticas (ej. cantidades, precios y valores) respecto a un momento o <u>período base</u>.

<u>Notación:</u>
p_0, q_0: precios o cantidades del año base
p_t, q_t: precios o cantidades del año *t*

Clasificación de los números índices

- Simples
- Complejos
  - Sin ponderar: media aritmética, media agregada.
  - Ponderados: i. de Laspeyres, i. de Paasche, i. de Fisher.

Índice simple
Estudia sólo una variable. Compara el valor actual con el del período base.
(Ver ejemplo en la siguiente tabla, de los índices complejos sin ponderar.)

$$I_0^t = \frac{x_t}{x_0} \times 100$$

ÍNDICES COMPLEJOS SIN PONDERAR

Los índices complejos tienen en cuenta más de una variable. Por ejemplo, la variación del precio de varios productos.

Índice de la media aritmética (Ima)
Para cada observación calculamos la media de los índices simples.

$$\text{Ima}_0^t = \frac{\sum_{i=1}^{M} I_0^{it}}{M}$$

Índice de la media agregada (Imag)
Para cada observación dividimos la suma de los valores actuales entre la suma de los valores en el período base.

$$\text{Imag}_0^t = \frac{\sum_{i=1}^{M} p_{it}}{\sum_{i=1}^{M} p_{i0}} \times 100$$

Tabla de índices simples e índices complejos sin ponderar

	Precios			Índices simples de precios			Índ. complejos	
Año	p_1	p_2	p_3	I_0^t (1)	I_0^t (2)	I_0^t (3)	Ima_0^t	Imag_0^t
2001	14	25	5,50	100,00	100,00	100,00	100,00	100,00
2002	15	26	5,75	107,14	104,00	104,55	105,23	105,06
2003	14	24	5,25	100,00	96,00	95,45	97,15	97,19
2004	16	27	6,00	114,29	108,00	109,09	110,46	110,11

Veamos, por ejemplo, cómo calcular los índices complejos del año 2004.

$$\text{Ima}_0^{2004} = \frac{114,29 + 108 + 109,09}{3} = 110,46$$

$$\text{Imag}_0^{2004} = \frac{16 + 27 + 6}{14 + 25 + 5,50} \times 100 = \frac{49}{44,5} \times 100 = 110,11$$

ÍNDICES COMPLEJOS PONDERADOS

Los índices complejos ponderados otorgan una importancia distinta a cada una de las variables, lo que se denomina “ponderación” (por ejemplo, dando más valor a un producto que se consume más, a la hora de calcular el índice de precios).

$$I = \sum_{i=1}^{M} I_{it} w_{it} \times 100$$

Nosotros calculamos los índices de Laspeyres y Paasche, de precios y de cantidades. Hay que tener en cuenta que:

- p_{i0} y q_{i0} se refieren a los precios y las cantidades del período base (que generalmente es el primer año, a no ser que se indique otra cosa).
- p_{it} y q_{it} se refieren a los precios y las cantidades del período actual.

Ejemplo: Ver tabla de la siguiente página. Para más detalles, consulta las diapositivas, la práctica 10 o el archivo “Ejemplo Número Índices.xls”.

Índice de Laspeyres de precios o de cantidades

$$L_{p_t} = \frac{\sum_{i=1}^{M} p_{it} q_{i0}}{\sum_{i=1}^{M} p_{i0} q_{i0}} \qquad L_{q_t} = \frac{\sum_{i=1}^{M} p_{i0} q_{it}}{\sum_{i=1}^{M} p_{i0} q_{i0}}$$

Índice de Paasche de precios o de cantidades

$$P_{p_t} = \frac{\sum_{i=1}^{M} p_{it} q_{it}}{\sum_{i=1}^{M} p_{i0} q_{it}} \qquad P_{q_t} = \frac{\sum_{i=1}^{M} p_{it} q_{it}}{\sum_{i=1}^{M} p_{it} q_{i0}}$$

Estos índices luego se multiplican por 100. (Está omitido para ahorrar espacio.)

Tabla de índices complejos ponderados

	Precios y cantidades						Índices complejos ponderados			
Año	p_1	q_1	p_2	q_2	p_3	q_3	L_{p_t}	L_{q_t}	P_{p_t}	P_{q_t}
2001	14	26	25	587	5,50	128	100,00	100,00	100,00	100,00
2002	15	30	26	621	5,75	135	104,10	106,00	104,10	106,01
2003	14	29	24	605	5,25	136	96,07	103,40	96,07	103,41
2004	16	33	27	650	6,00	143	108,19	111,15	108,21	111,17

Veamos, por ejemplo, cómo calcular los índices complejos ponderados del año 2003.

$$L_{p_{2003}} = \frac{14\cdot 26+24\cdot 587+5,25\cdot 128}{14\cdot 26+25\cdot 587+5,50\cdot 128}\times 100 = \frac{15124}{15743}\times 100 = 96,07$$

$$L_{q_{2003}} = \frac{14\cdot 29+25\cdot 605+5,50\cdot 136}{14\cdot 26+25\cdot 587+5,50\cdot 128}\times 100 = \frac{16279}{15743}\times 100 = 103,40$$

$$P_{p_{2003}} = \frac{14\cdot 29+24\cdot 605+5,25\cdot 136}{14\cdot 29+25\cdot 605+5,50\cdot 136}\times 100 = \frac{15640}{16279}\times 100 = 96,07$$

$$P_{q_{2003}} = \frac{14\cdot 29+24\cdot 605+5,25\cdot 136}{14\cdot 26+24\cdot 587+5,25\cdot 128}\times 100 = \frac{15640}{15124}\times 100 = 103,41$$

<u>TEMA 7:</u> ANÁLISIS ESTADÍSTICO DE DATOS ORDINALES Y CATEGÓRICOS

En este tema estudiaremos la relación entre varias variables que no son cuantitativas, sino cualitativas. Se trata de calcular una medida similar a la correlación.

Estas variables las llamamos <u>atributos</u>. Pueden estar en una escala ordinal o nominal, según tengan o no un orden, respectivamente.

ANÁLISIS DE DATOS ORDINALES: CORRELACIÓN POR RANGOS

En este caso estudiamos la relación que existe entre dos variables, a partir del orden que ocupa cada observación. Este orden se denota como x_i e y_i.

Para calcular esta relación utilizamos el siguiente método.

Coeficiente de correlación por rangos de Spearman

$$\rho = 1 - \frac{6 \cdot \sum_{i=1}^{N} \left(x_i - y_i \right)^2}{N^3 - N}$$

El resultado de este coeficiente está acotado entre –1 y 1. Se interpreta así:

- Si $\rho = 0$, no existe correlación.
- Si $\rho = 1$, existe concordancia perfecta.
- Si $\rho = -1$, existe disconcordancia perfecta.

<u>Ejemplo:</u>

Estudiar la relación entre las notas de matemáticas y estadística, a partir del orden que ocupa cada observación.

Mates	Estadística	X (orden mat.)	Y (orden est.)	$(X - Y)^2$
10	8	1	4,5	12,25
9	6	2	7,5	30,25
8	10	3	1	4
7	9	4,5	2,5	4
7	8	4,5	4,5	0
6	7	6,5	6	0,25
6	6	6,5	7,5	1
4	9	8	2,5	30,25

Primero sumamos la última columna: $\sum_{i=1}^{N} \left(x_i - y_i \right)^2 = 82$

Ahora aplicamos la fórmula del coeficiente de correlación de Spearman:

$$\rho = 1 - \frac{6 \cdot \sum_{i=1}^{N} \left(x_i - y_i \right)^2}{N^3 - N} = 1 - \frac{6 \cdot 82}{8^3 - 8} = 1 - \frac{492}{504} = 0,024$$

Las dos variables están muy poco relacionadas, dado que ρ es próximo a 0.

ANÁLISIS DE DATOS NOMINALES

Coeficiente de asociación H

Podemos usarlo cuando sólo tenemos 2 modalidades para cada variable (tablas 2×2).

Los datos se colocan en una tabla de contingencia como la que se muestra a continuación. En este caso se estudia la relación entre ser hombre/mujer y fumar.

	Hombre	Mujer	Marginal
Fuma	n_{11}	n_{12}	$n_{1\cdot}$
No fuma	n_{21}	n_{22}	$n_{2\cdot}$
Marginal	$n_{\cdot 1}$	$n_{\cdot 2}$	N

Se utiliza la siguiente fórmula, sustituyendo los valores a partir de la tabla.

$$H = n_{11} - \frac{n_{1\cdot}n_{\cdot 1}}{N} = \frac{Nn_{11} - n_{1\cdot}n_{\cdot 1}}{N}$$

Interpretación:

- Si H = 0, los atributos son independientes.
- Si H > 0, presentan una asociación positiva.
- Si H < 0, presentan una asociación negativa.

Inconvenientes: no está acotado (difícil de interpretar), y sólo sirve para tablas 2×2.

Tablas de contingencia *h* × *k*

En este caso estudiaremos la relación entre 2 atributos (variables cualitativas).

- *h* es el número de modalidades (posibles valores) del atributo A.
- *k* es el número de modalidades (posibles valores) del atributo B.

Para estudiar la relación, hay que construir una tabla de valores esperados, que son los valores que ocurrirían si los atributos fueran independientes.

La tabla de valores esperados se rellena con esta fórmula: $n'_{ij} = \frac{n_{i\cdot}n_{\cdot j}}{N}$

(Para cada celda se multiplican las distribuciones marginales y se divide entre el total.)

Coeficiente de contingencia χ^2:

Para estudiar la relación se calcula χ^2 usando la tabla original y la de valores esperados.

Para obtener χ^2, en cada celda se calcula el cuadrado de la diferencia de la celda de la tabla original (n_{ij}) respecto a la misma celda de la tabla de valores esperados (n'_{ij}), y se divide entre esta última celda (n'_{ij}). Después se suman todos esos valores.

$$\chi^2 = \sum_{i=1}^{h}\sum_{j=1}^{k}\frac{\left(n_{ij} - n'_{ij}\right)^2}{n'_{ij}}$$

Coeficiente de Tschuprow:

Permite interpretar la relación entre los atributos, ya que está acotado entre 0 y 1.

$$T^2 = \frac{\chi^2/N}{\sqrt{(h-1)(k-1)}}$$

Si $T^2 = 0$, los atributos son independientes. Si $T^2 = 1$, están totalmente asociados.

Ejemplo:

Tabla de la relación entre el nivel de colesterol y la zona de residencia (costa/interior):

	Nivel de colesterol			
	Bajo	Medio	Alto	Marginal
Zona costa	15	18	7	40
Z. interior	2	22	23	47
Marginal	17	40	30	87

Estudiaremos la relación entre los 2 atributos calculando el coeficiente de Tschuprow. Primero construimos la tabla de valores esperados:

	Nivel de colesterol			
	Bajo	Medio	Alto	Marginal
Zona costa	7,8	18,3	13,8	40
Z. interior	9,18	21,6	16,2	47
Marginal	17	40	30	87

¿Cómo se calculó? Veamos 2 ejemplos: $n'_{11} = \frac{40 \cdot 17}{87} = 7{,}8 \quad n'_{21} = \frac{47 \cdot 17}{87} = 9{,}18$

Ahora obtenemos χ^2 y después T^2.

$$\chi^2 = \frac{(15-7{,}8)^2}{7{,}8} + \frac{(18-18{,}3)^2}{18{,}3} + \frac{(7-13{,}8)^2}{13{,}8} + \frac{(2-9{,}2)^2}{9{,}2} + \frac{(22-21{,}6)^2}{21{,}6} + \frac{(23-16{,}2)^2}{16{,}2} = 18{,}35$$

$$T^2 = \frac{\chi^2/N}{\sqrt{(h-1)(k-1)}} = \frac{18{,}35/87}{\sqrt{(3-1)(2-1)}} = \frac{18{,}35/87}{\sqrt{2}} = 0{,}149$$

Los atributos están poco relacionados, ya que T^2 es próximo a 0.